

# 基于建模的 NoC 性能评价方法研究

韩立敏 任向隆 高德远 樊晓桢 安建峰

(西北工业大学计算机学院 西安 710072)

**摘要** NoC 的设计空间非常庞大,并且不同方案间存在巨大的性能差异,所以在设计初期对不同的设计方案进行性能评价极其重要。以基于建模的 NoC 性能评价方法为主线,重点分析和对比基于仿真模型和基于分析模型的性能评价方法的研究现状和评价能力,由此归纳出 NoC 评价模型的准确度评估方法。最后,在对面临的挑战进行分析的基础上,提出基于建模的 NoC 性能评价的发展方向。

**关键词** 片上网络,性能评价,仿真模型,分析模型

**中图分类号** TP302 **文献标识码** A

## Modeling-based Performance Evaluation Methodology of Network on Chip

HAN Li-min REN Xiang-long GAO De-yuan FAN Xiao-ya AN Jian-feng

(Department of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China)

**Abstract** The entire design space is extremely large, and there is huge difference of performance between different options. So performance evaluation of different design becomes an important issue in initial stages. The paper took modeling-based performance evaluation methodology as masterstroke, discussed research status and metrics of evaluation methods which include simulation modeling and analysis modeling, and summarized accuracy estimating methods of NoC evaluation models. Also, a comparison was made between methods mentioned above. Finally, a guidance of future research was proposed based on the analysis of future challenges.

**Keywords** Network on chip, Performance evaluation, Simulation model, Analysis model

## 1 引言

随着半导体工艺的迅猛发展,单芯片的集成能力越来越强,以 AMBA、Wishbone 为代表的传统总线的信号延迟与功耗急剧增加,导致其可扩展性差,不再适用于未来片上众核互连的通信要求。近年来,学术界与工业界借鉴网络领域的相关概念提出了一种新的通信技术——片上网络(Network on Chip, NoC)<sup>[1]</sup>,用于支持未来众核处理器中处理器核心之间的互连通信。

NoC 结构与传统总线相比,具有可扩展性好、能量效率高、可重用性高等特点,是未来片上众核处理器间实现核间互联的首选方案。然而, NoC 也暴露出设计复杂性问题:在设计初期,拓扑结构、传输协议、路由器流水线级数、缓冲数量等各个参数均存在多种选择,导致设计空间庞大,并且不同方案之间存在巨大的性能差异。因此,在系统设计初期,探索一种 NoC 的性能评价方法显得十分必要;另外,较为精确的网络性能评价,对早期方案制定、产品及早上市都具有重要意义。

NoC 的性能评价方法大体分为 3 类:基于测量的性能评

价、基于仿真模型的性能评价和基于分析模型的性能评价。基于测量的性能评价需在构建完毕的 NoC 中加入相应的测量模块,之后通过运行系统获取系统的性能参数,因此,这种方法不能用于 NoC 设计初期设计空间的探索、方案的制定;而基于建模的性能评价包括基于仿真模型的性能评价和基于分析模型的性能评价。它们可用于 NoC 的设计和开发阶段,而评测处于设计阶段的 NoC 的性能。鉴于上述原因,本文不对基于测量的方法进行研究,而仅讨论基于建模的 NoC 性能评价方法。

本文第 2 节介绍基于仿真模型的性能评价方法;第 3 节介绍和分析基于分析模型的性能评价方法;第 4 节介绍和分析 NoC 评价模型的准确度评估方法;第 5 节将第 2 节和第 3 节的评价方法进行比较;最后为结束语,在此指出基于建模的 NoC 性能评价面临的挑战及发展方向。

## 2 基于仿真模型的性能评价

对 NoC 的资源及资源间的连接使用建模语言进行描述可以构建仿真模型,并在该模型上运行测试程序能达到评价

到稿日期:2012-01-05 返修日期:2012-05-01 本文受国家自然科学基金(61173047, 61003037, 60736012), 国家“863”基金项目(2009 AA01Z110)和西北工业大学基础研究基金“性能约束的片上网络低功耗设计技术”资助。

韩立敏(1983—),女,博士生,主要研究领域为计算机体系结构、片上存储系统, E-mail: hanlm@mail.nwpu.edu.cn;任向隆(1982—),男,博士生,主要研究领域为计算机体系结构、片上网络;高德远(1946—),男,教授,博士生导师,主要研究领域为高性能处理器体系结构;樊晓桢(1962—),教授,博士生导师,主要研究领域为高性能处理器体系结构、VLSI 设计及计算机网络;安建峰(1977—),男,博士,讲师,主要研究领域为计算机体系结构、数字系统设计。

NoC 性能的目的。基于仿真模型的 NoC 性能评价通常的做法是:选择一个自有的或其他研究机构开发的 NoC 仿真器;根据设计需求对 NoC 的相关参数进行选择,得到一种特定的配置;使用该配置在所选的仿真器上运行测试程序,得到所关心的 NoC 性能评价指标。

在 NoC 的概念被提出后很快出现了许多软件仿真器。如瑞典皇家理工学院 Nostrum 项目组开发的 NNSE<sup>[2]</sup>, 在一定的网络配置和流量模式下,可得到网络的吞吐量和延迟;美国普林斯顿大学开发的 LUNA<sup>[3]</sup>, 是一款周期精确的链路功耗仿真器;而其后开发的另一款 NoC 功耗、性能仿真器 Orion<sup>[4]</sup>, 可用于动态功耗、漏电流及性能分析;另外,其随后开发了详细的 NoC 模型 Garnet<sup>[5]</sup>, 并集成到了当前较为流行的全系统仿真器 GEMS<sup>[6]</sup> 中;此外,意大利卡塔尼亚大学计算机体系结构小组开发的 Noxim<sup>[7]</sup> 和英国南安普敦和印度马拉雅亚国立理工学院联合开发的 Nirgam<sup>[8]</sup> 等软件仿真器,能对延迟、吞吐量以及功耗等 NoC 参数进行评估。

这些仿真器大多使用 SystemC 开发,例如 NNSE、Noxim 和 Nirgam。表 1 给出了通常的输入、输出参数,其中输入参数分为两类:NoC 相关参数和应用相关参数。NoC 相关参数对片上网络结构本身进行配置,应用相关参数确定在该结构上运行的应用特征。根据这些输入,仿真器可得到相应的输出,即 NoC 性能评价指标。此外,最新的 Noxim 仿真器包含了自己的功耗模型,除表 1 提到的评价指标外,它还可以对全局的总能耗、每次通信的能耗进行统计;最新的 Garnet 和 Nirgam 仿真器集成了普林斯顿的 Orion 模型,可对路由器的功耗(包括动态功耗和漏电流)进行估计,除功耗外,Nirgam 还可对路由器的面积进行估计。

表 1 软件仿真器通常的输入和输出参数

类别	具体参数
输入参数	NoC 相关 拓扑结构、交换技术、路由机制、链路带宽、流控方法、是否支持 QoS 等
	应用相关 网络负载、应用程序映射、流量模式等
输出参数	接收到的总包/切片数、最大/小全局延时、每次通信的延时/吞吐量、全局平均延时/吞吐量等

### 3 基于分析模型的 NoC 性能评价

基于分析模型的 NoC 性能评价方法对 NoC 的资源及资源间的连接使用数学工具进行建模,通过估计资源利用率以及时间方面的特性来预测 NoC 的性能。在 NoC 的性能分析建模中,排队论<sup>[9]</sup>是必不可少的数学工具。

之前,大多数网络性能分析工作均集中在多处理器与宏观网络通信领域。如,Dally 等人<sup>[10]</sup>最早借鉴 M/M/1 排队系统分析了 k 维 n 立方体网络中的传输延迟;随后,文献[11-13]采用不同方法来评估均匀传输分布下的宏观网络性能。这些研究要么在高流量负载下的分析误差较大,要么仅针对均匀分布传输,并假设无穷缓冲或者单切片通道,与 NoC 的实际情况不符,因而这些分析模型不能在 NoC 中使用。

近几年,针对 NoC 的性能分析工作才逐渐展开。Hu 等人<sup>[14]</sup>最早借助 M/M/1/K 建立了一种 NoC 解析模型,其目的在于针对目标程序来优化缓冲资源。该方法节省了 80% 的缓冲,并且没有任何性能损失,但不足的是它仅局限于固定

的片上互连结构,并且只能采用存储转发或虚拟直通交换策略。随后,Ogras 等人<sup>[15]</sup>提出了采用虫孔交换策略的路由器模型,它属于粗粒度的数学建模,并且忽视了路由器实现细节,在高流量下分析精度不足,无法指导路由器结构设计。与此同时,Huang 等人<sup>[16]</sup>也对片上互连结构进行了建模,主要讨论如何分配传输通道问题。该模型缺乏对通信结构中各种阻塞现象的分析,所以无法获得吞吐量、传输延迟等通信性能参数。此后,Lai 等人<sup>[17]</sup>从典型 NoC 路由器结构出发,分析报文传输中的各种阻塞现象,建立了基于 M/G/1/N 排队系统的路由器解析模型,并提出了网络性能分析算法。该方法获得了较高的分析精度,同时使分析效率提高了 200 倍左右,适合进行 NoC 网络性能评估,并有助于指导程序拓扑映射、分析性能瓶颈等关键工作。

NoC 的分析模型大多基于排队系统进行建模,并且对建模对象进行简化,对网络拓扑、路由算法、流控策略等进行指定,例如,通常假定为 2D Mesh 结构、XY 确定路由。分析模型的输入参数主要有两个:包/切片的到达速率、路由器的服务率。这两个参数可根据 NoC 的拓扑、路由算法、流控策略和应用的流量特征等分析得到。有这两个参数后,可根据排队系统和 Little 公式求得平均等待队列长度、平均系统队列长度、平均队列等待时间、平均系统等待时间等输出参数,如表 2 所列。依据这些输出参数,可进一步得到缓存/虚链路利用率、平均延迟等网络性能参数,据此可对 NoC 的性能进行评价和优化。

表 2 分析模型输入和输出参数

类别	具体参数
输入参数	包/切片的到达速率、路由器的服务率
输出参数	平均等待队列长度、平均系统队列长度、平均队列等待时间、平均系统等待时间

### 4 NoC 评价模型的准确度评估

无论是基于仿真模型,还是基于分析模型的性能评价方法,研究人员在开发出一款性能评价模型后,都需要对模型的准确度进行评估,以便既可验证自己的工作,也可以辅助其他研究人员在选择该模型 NoC 做方案决策时做到心中有数。通常,对模型进行准确度评估的方法有以下两种:

方法 1 将评价模型的结果与其他已公开发表、得到公认的性能评价模型的结果对比,并给出偏差范围;

方法 2 将评价模型的结果与低层次(如 IC 后端)设计工具的评估结果对比,并给出偏差范围。

表 3 列出了普林斯顿大学开发的 Garnet 仿真模型、Orion 仿真模型和 Lai 等人提出的分析模型所采用的模型准确度评估方法和比较对象。

表 3 模型准确度评估方法举例

模型名称	评估方法	比较对象
Garnet 仿真模型	方法 1	与文献[18,19]中的结果进行了对比,此外,还与 PoPNet <sup>[20]</sup> 仿真器的结果进行了对比
Orion 仿真模型	方法 2	与文献[21]中给出的 Intel 80-Core 布局布线后的仿真结果和文献[22]给出的 Intel SCC 布局布线前的仿真结果进行对比
Lai 等人 <sup>[17,23]</sup> 提出的分析模型	方法 1	与 Nirgam 仿真器的结果进行对比

## 5 两种评价方法的比较

在 NoC 性能评价方面,仿真是一种用途极为广泛并且有用的工具。不像分析模型那样,其所能建模的特征范围受到限制,仿真模型几乎可在任意细节水平上对目标进行构建。所以相对于分析模型来说,在极为棘手的复杂情形,仿真模型依然可以采用仿真方法进行建模。另外,仿真模型可对动态或瞬态的行为进行研究,可以提供特征的分布估计和瞬时值,而分析模型通常仅能用来研究稳态行为,只能提供特征的平均值(见表 4)。事实上,仿真的一个主要的应用就是对分析模型进行验证。而仿真方法的不足在于,每次仿真耗时长,不适合大尺度设计空间的探索,仅能选择有限的设计方案进行评估,影响了优化效果,且浪费了大量的资源,给产品的研发周期造成了负面影响。

表 4 两种评价方法的比较

类别	基于仿真模型的性能评价	基于分析模型的性能评价
模型细节	任意水平	细节有限
拓扑结构	多种选择	固定
交换方式	多种选择	固定
路由算法	多种选择	固定
自定义路由	大多支持	不支持
自适应路由	易于实现	无法扩展
流控策略	多种选择	固定
总包/切片数	支持	不支持
每次通信延时/吞吐量	支持	不支持
最大/小延时/吞吐量	支持	不支持
平均延时/吞吐量	支持	支持
耗时	长	短
大规模设计空间探索	不适合	适合
灵活性	好	差

分析模型作为一种代价有效的性能评估技术已经被广泛地接受。分析模型之所以是代价有效的,是因为它基于数学等式的有效求解。然而,为使这些等式有一个易于得到的解,在考虑排队论模型的结构和行为时,必须进行一些假定,以进行简化。分析模型的代价有效性体现在其速度很快,适合对大尺度的设计空间进行探索,从而得到较好的优化效果,但又因其进行了一些简化的假定,分析模型不可能捕捉到在仿真模型中构建的所有细节。尽管如此,对许多应用来说,分析模型仍足以对关键资源和负载需求进行建模,找到影响系统性能的瓶颈和关键参数。

以文献[17]选取的应用为例,对于  $4 \times 4$  的 2D Mesh 网络,探索整个设计空间,仿真方法预计耗时 4368h,而分析方法耗时约 23h,效率提高了约 200 倍;对于现有的 NoC 仿真器<sup>[2,5,7,8]</sup>,拓扑结构、路由算法等均有多种选择,如拓扑结构通常有 Mesh、Torus、Fat Tree 等,路由算法通常有 XY、OE 等,有的仿真器还支持用户自定义拓扑结构和路由算法,而现有的分析模型<sup>[14-17,23,24]</sup>一般都假定拓扑为 Mesh 结构、路由策略为 XY 确定路由;另外,由于分析模型需要预先确定的排队系统参数,因此其路由策略不易扩展到自适应路由,而对于仿真模型,则较为容易。

**结束语** 随着技术的发展, NoC 将更为复杂、规模将更大,设计空间膨胀对性能评价方法提出了更高的要求,如何快速有效地得到具有较好效果的优化方案将是设计人员面临的

一个重要问题。随着系统复杂度的提高,分析模型的构建变得更为困难,在 NoC 性能评价方面仿真的重要性将进一步增加。然而,如第 5 节所述,仿真具有较大的时间开销,特别是多核、众核处理器架构的出现,使得基于仿真模型的评价方法很难在时钟精确级别进行,也很难在有限的时间内对大规模设计空间进行有效的探索。

在总结和分析现有性能评价方法存在的问题和面临挑战的基础上,本文认为未来的工作可在以下几个方面展开:

(1) 仿真技术。研究如何减少仿真器实际运行的负载,包括减少标准性能测试程序的参数输入集和减少仿真运行的指令数;研究如何利用多核处理器、多核 GPU 加速仿真的运行速度,通过对现有 NoC 仿真器的修改,使其在多核处理器、多核 GPU 上高效运行,甚至可以将分布式仿真技术应用到 NoC 的仿真中。

(2) 分析技术。研究新的数学理论和在现有理论基础上对 NoC 进一步建模,使分析建模技术始终紧跟 NoC 技术的发展。

(3) 层次建模技术。在仿真建模和分析建模中应用层次建模技术。根据不同的需求,在不同层次上对 NoC 进行建模,获得不同的细节水平和评估代价。进行性能评价时,根据需要选择其中的一种或几种模型层次,以在合理评估代价的约束下,获得较好的评估细节。

(4) 混合建模技术。其包含仿真和分析模型的混合建模技术,这两种模型形成互补,以获得更好的效果。例如,使用简化的分析模型对 NoC 的设计空间进行探索,得到一组优化方案集合,再利用仿真模型进行细粒度的仿真,从这组优化方案中挑选出最优方案。

## 参考文献

- [1] Dally W J, Towles B. Route packets, not wires, on-chip interconnection networks[C]// Design Automation Conference. 2001: 683-689
- [2] Lu Zhong-hai, Thid R, Millberg M, et al. NNSE: Nostrum network-on-chip simulation environment[C]// Swedish System-on-Chip Conference. 2005: 1-4
- [3] Eisley N, Li-Shiuan P. High-level power analysis for on-chip networks[C]// CASES'04. 2004: 104-115
- [4] Wang Hang-sheng, Zhu Xin-ping, Li-Shiuan P, et al. Orion: A power-performance simulator for interconnection networks[C]// MICRO'35. 2002: 294-305
- [5] Agarwal N, Krishna T, Li-Shiuan P, et al. GARNET: A detailed on-chip network model inside a full-system simulator[C]// IEEE ISPASS. 2009: 33-42
- [6] Martin M M, Sorin D J, Beckmann B M, et al. Multifacet's general execution-driven multiprocessor simulator (GEMS) toolset[J]. ACM SIGARCH Computer Architecture News, 2005, 33(4): 92-99
- [7] Palesi M, Patti D, Fazzino F. Noxim: the noc simulator user guide [EB/OL]. [http://noxim.sourceforge.net/pub/Noxim\\_User\\_Guide.pdf](http://noxim.sourceforge.net/pub/Noxim_User_Guide.pdf), 2011-02-23
- [8] Nirgam J L. A simulator for noc interconnect routing and applications modeling, v1. 1 [EB/OL]. <http://nirgam.eecs.soton.ac.uk/Documentation.php>, 2011-02-23

- [9] 陆传贵. 排队论(第二版)[M]. 北京:北京邮电大学出版社, 2009:31-38
- [10] Dally W J. Performance analysis of k-ary n-cube interconnection networks[J]. IEEE Transactions on Computers, 1990, 39(6): 775-785
- [11] Adve V S, Vernon M K. Performance analysis of mesh interconnection networks with deterministic routing[J]. IEEE Transactions on Parallel and Distributed Systems, 1994, 5(3): 225-246
- [12] Agarwal A. Limits on interconnection network performance[J]. IEEE Transactions on Parallel and Distributed Systems, 1991, 2(4): 398-412
- [13] 侯国峰, 杨愚鲁. 超级递归基准互连网络性能分析[J]. 计算机科学, 2001, 28(10): 85-88
- [14] Hu Jing-cao, Ogras U Y, Marculescu R. System-level buffer allocation for application-specific networks-on-chip router design [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006, 25(12): 2919-2933
- [15] Ogras U Y, Marculescu R. Analytical router modeling for networks-on-chip performance analysis[C]// DATE'07. 2007: 1-6
- [16] Huang Ting-chun, Ogras U Y, Marculescu R. Virtual channels planning for networks-on-chip[C]// ISQED'07. 2007: 879-884
- [17] 赖明澈, 王志英, 戴葵. 基于路由器解析式模型的 NoC 网络性能分析方法[J]. 计算机辅助设计与图形学学报, 2009, 21(3): 339-345
- [18] Kumar A, Li-Shiuan P, Kundu P, et al. Express virtual channels; towards the ideal interconnection fabric[C]// ISCA'07. 2007: 150-161
- [19] Nicopoulos C A, Park D, Kim J, et al. Vichar: a dynamic virtual channel regulator for network-on-chip routers [C]// MICRO' 39. 2006: 333-346
- [20] PoPNet[EB/OL]. <http://www.princeton.edu/edu/?lshang/popnet.html>
- [21] Hoskote Y, Vangal S, SINGH A, et al. A 5-GHz Mesh Interconnect for a Teraflops Processor[J]. Micro, IEEE, 2007, 27(5): 51-61
- [22] Llitzky D A, Hoffman J D, Chun A, et al. Architecture of the Scalable Communications Core's Network on Chip[J]. Micro, IEEE, 2007, 27(5): 62-74
- [23] Lai Ming-che, Gao Lei, Xiao Nong, et al. An accurate and efficient performance analysis approach based on queuing model for network on chip[C]//IEEE/ACM ICCAD. 2009: 563-570
- [24] Hu Jing-cao, Marculescu R. Application-specific buffer space allocation for networks-on-chip router design [C]// IEEE/ACM ICCAD. 2004: 354-361

(上接第 266 页)

- [2] 赵辉, 陈辉, 于泓. 一种改进的全景图自动拼接算法[J]. 中国图像图形学报, 2007, 2(12): 336-342
- [3] 冈萨雷斯. 数字图像处理[M]. 北京: 电子工业出版社, 2009
- [4] Szeliski R. Image alignment and stitching, A tutorial [J]. Foundations and Trends in Computer Graphics and Vision, 2006, 2(1): 1-104
- [5] 王娟, 师军, 吴宪祥. 图像拼接技术综述[J]. 计算机应用技术研究, 2008, 5(7): 1940-1947
- [6] 章毓晋. 图像工程: 图像处理和分折[M]. 北京: 清华大学出版社, 1999
- [7] Chalechale A, Naghdy G, Mertins A. Sketch based image matching using angular Partitioning systems [J]. IEEE Trans on Man and Cybernetics, Part A, 2005, 35(1): 28-41
- [8] Zhu Q, Wu B, Xu Z. Seed point selection method for triangle const rained image matching propagation [J]. IEEE Geoscience and Remote Sensing Letters, 2006, 3(2): 207-211
- [9] Zomet A, Levin A, Peleg S, et al. Seamless image stitching by minimizing false edges [J]. IEEE Trans on Image Processing, 2006, 15(4): 969-977
- [10] 涂春萍, 柴亚辉, 李广丽, 等. 一种基于 Harris 角点特征精确匹配的图像拼接方法[J]. 实验室研究与探索, 2011, 30(10): 40
- [11] 庄志国, 孙惠军, 董继扬, 等. 基于角点检测的图像匹配算法及其在图像拼接中的应用[J]. 厦门大学学报, 2007, 46(4): 501-505
- [12] 宋宝森, 付永庆, 宋海亮. 一种消除图像拼接的新方法[J]. 计算机科学, 2012, 38(2): 260-263

(上接第 285 页)

## 参 考 文 献

- [1] Borges P V K, Izquierdo E. A probabilistic approach for vision-based fire detection in videos[J]. IEEE Transaction on Circuits and Systems for Video Technology, 2010, 20(5): 721-731
- [2] Liu Che-bin, Ahuja N. Vision based fire detection [C]// Proceedings of International Conference on Pattern Recognition. 2004: 134-137
- [3] 张正荣, 李荣刚. 基于支持向量机的火灾探测技术[J]. 微型机与应用, 2010, 29(24): 70-72
- [4] 杨国田, 吴章宪, 杨鹏远. Boosting 在火灾识别中的应用研究 [J]. 计算机工程与应用, 2010, 46(5): 200-204
- [5] Grimson W E L, Stauffer C, Romano R, et al. Using adaptive tracking to classify and monitor activities in a site[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: 1998: 22-29
- [6] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]// Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Fort Collins, Colorado: 1999
- [7] Stauffer C. Learning patterns of activity using real-time tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757
- [8] Power P W, Schoonees J A. Understanding background mixture models for foreground segmentation[C]// Proceeding of Image and Vision Computing. New Zealand, 2002: 267-271